REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
21-10-2011	Final Report	1-2-2008-30-6-2011

4. TITLE AND SUBTITLE

Distributed Information Processing for Battlespace Awareness: Ergodic and Non-Ergodic Interplay

5a. CONTRACT NUMBER
FA9550-08-1-0060

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Aaron B. Wagner

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cornell University Sponsored Program Services 120 Day Hall Ithaca, NY 14853

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research 875 North Randolph Street Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

AFRL-OSR-VA-TR-2012-0818

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The core of this project was devoted to distributed compression of multimodel (vector) sources and distributed compression for hypothesis testing.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Aaron Wagner
public	public	public		2	**19b. TELEPHONE NUMBER** *(include area code)* 607-255-1017

Final Report for FA9550-08-1-0060
DISTRIBUTED INFORMATION PROCESSING FOR BATTLESPACE AWARENESS---ERGODIC AND NON-ERGODIC INTERPLAY

Aaron Wagner
School of Electrical and Computer Engineering
Cornell University

Per the original proposal, the core of this project was devoted to distributed compression of multimodel (vector) sources and distributed compression for hypothesis testing.

In the final months of the project, we **cracked** the problem of determining the rate region of the vector Gaussian "one-helper" source coding problem. This problem was **one of the most fundamental open problems** in information theory, and had withstood repeated attacks by several groups around the world, starting with Liu and Viswanath (2007). The problem is similar to that of determining the capacity region of the Gaussian MIMO broadcast channel, whose solution won two awards from the IEEE Information Theory Society, but the compression version of the problem turned out to be significantly harder. Our proof technique used the method introduced to solve that problem but also used a fundamentally new technique that we call "distortion projection," which essentially involves **projecting the problem into a lower-dimensional space** where it is easier to analyze. Our results imply that a very simple compression algorithm is optimal for this problem. **The PI considers this result to be the best result to come out of his group in the last 5 years, among all projects.**

We also showed that for the discrete memoryless version of this problem, we have shown that the existing state-of-the-art compression scheme is suboptimal. **This also resolved an open problem in the literature, this one being a 30-year-old open problem in network information theory.** To show this, we introduced a new compression scheme based on isolating common components that performs strictly better than existing schemes. We also showed that this new scheme is optimal in certain cases.

Distributed compression for hypothesis testing is a fundamental---and extremely challenging---problem that arises in many application areas including traffic analysis in networks, radar systems, wireless relays, and sensor networks. Yet despite the fundamental nature of this problem, relatively little was known about it. In particular, it was not known how to optimally compress data when the goal is not to reproduce it at the destination but instead to make an inference. It was not even known if binning, a commonly-used primitive in distributed compression, should be used in this scenario: while binning leads to increased compression ratios, in some cases its failure rate dominates the overall system performance. We showed that binning-based compression schemes are actually optimal for a class of distributed inference problems. This shows that, from a compression standpoint, binning is effective for inference even though, somewhat paradoxically, its errors may dominate system performance. We then used this result to show exhibit a compression scheme that is nearly optimal for a much wider class of problems. This result has drawn renewed attention to an important but neglected area, and other

researchers are examining how our techniques can be applied in application-specific areas such as wireless relaying.

During the early stages of the project, we leveraged the connection between Gaussian and discrete erasure problems to develop a much better understanding of erasure compression problems. We answered the following question: suppose that any k out of n packets are enough to recover the source, how much of the source can we recover with 1, 2, … , k - 1 packets? **That is, how many bits can I decode as a function of the number of bits I have received?** Existing schemes exhibit a "cliff" effect: one cannot decode any of the source until one can decode all of it. We developed simple schemes that allow one to decode more of the source the more packets one receives. Moreover, these simple scheme is provably optimal. These results required inventing new analysis techniques that are suited to erasure problems.

We also initiated work in studying coding schemes for secure free-space **quantum-optical** and **timing-based** communication. Existing studies of information-theoretic security in wireless systems focuses mainly in RF-based systems. Accurate channel state information (CSI) is difficult to obtain for these systems, however, due to small-scale fading, and existing results have very strigent CSI requirements. Free-space optical communication is less prone to small-scale fading, which makes it much more amenable to information-theoretic security guarantees. We have characterized the secrecy capacity of the Poisson channel model of free-space optical channels and gave an explicit characterization of codes that achieves this capacity. The converse proving technique is novel and can be applied to other large-bandwidth channels. We also determined the capacity of the single-server, memoryless queue as a model for impulse-radio systems. This settled an open problem in the information theory literature and had several interesting implications. We showed that slow memoryless queues are stochastically degraded with respect to faster ones, which implies in particular that slow queues are more entropy increasing than faster ones. This **strengthened a result of Prabhakar and Gallager** (2003) that said that memoryless queues are entropy increasing.

Existing analyses of the performance limits of channel codes are based on large-deviations or central limit theorem asymptotics. We have shown that the moderation deviations asymptotic that has been introduced in probability theory has much more engineering relevance in the channel coding context, and we exactly determined the best possible moderate-deviations performance of codes. We also derived new sharper bounds in the large-deviations regime. These essentially determine the order of the pre-factor of the error exponent at rates close to capacity. **These bounds significantly improve upon the state of the art bounds, which were due to Shannon, Gallager, and Berlekamp in 1967.**

We have also designed codes for peer-to-peer networks subject to "pollution" attacks, i.e., subject to the possibility that adversaries can maliciously inject arbitrary packets into the network. We found optimal codes for this problem for Gaussian sources subject to MSE distortion, binary sources subject to Hamming distortion, and binary sources subject to erasure distortion. For erasure distortion, we found that separate source and channel coding is not optimal: the optimum strategy is to mix the two. While there exist instances for which separation is known to be suboptimal, the reason that separation fails here seems to be fundamentally different from these standard examples.